International Journal of Engineering & Technology, 7 (2.22) (2018) 53-56

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

RankTagViz: A Semantic Ranking and Tags Visualization of User Travelogues

M. Uma Maheswari[1*], Dr. J.G.R. Sathiaseelan[2]

[1]*Research Scholar, Department of Computer Science, Bishop Heber College, Tiruchirappalli, TN, India*
[2]*Head, Department of Computer Science, Bishop Heber College, Tiruchirappalli, TN, India*
**Email: umamanohar89@bhc.edu.in*

Abstract

The travelogues contain prosperous geo-referenced information such as tours, weather, and expenses etc. Reading travelogues and finding useful trip information is a tedious task for many people. So as to correlate the georeferenced text meaningfully the semantic ranking is taken in account on the user contributing travelogues. In this way there is a requirement for creating programmed travelogue mining methods to pass on valuable data in a travelogue to its peruses in the more successful way. Visualizing a geo-referenced data with location tags and descriptions makes it helpful for readers to understand the fundamental substance of the travelogue. For user expedient, this paper proposes a Travelogue RankTagViz approach that semantically ranks and visualizes the tags of user travelogues. The RankTagViz contains two phases. In the first phase, the user travelogues are ranked based on semantic. In the next phase, the travelogues are visualized based on the tag and images. During semantic ranking, a semantic dimension reduction method is proposed to pre-process and mine useful information from the travelogues. After that a semantic rank mechanism is proposed to rank travelogues based on tags and POI (Point Of Interest). For tag visualization, location based tags and images are extracted from the travelogues and a novel UI is intended to give a superior client encounter, by arranging both the literary and visual data produced by committed voyagers in an alluring way. Test comes about on an arrangement of gathered travelogues show the proposed techniques' capacity to rank and imagine travelogues.

Keywords: *Travelogues, Semantic, Ranking, Travel Tags, Visualization*

1. Introduction

Travel, as a necessary piece of mankind's history, has turned out to be increasingly mainstream in individuals' regular day to day existences lately, mostly inferable from the expanding measure of movement related data and administrations on the Web, which give individuals productive approaches to online arrangement and plan for their excursions. In the tourism space, people have capacity to record and offer their movement encounters on weblogs, dialog or travel groups, as scholarly travelogues and photographs taken amid the treks. Travelogues for the most part contain rich travel data, for example, the tours, accommodation, meals, and expenses, weather conditions, and so on, are exceedingly valuable to a trip planner. Thus, many individuals turn to on-line travelogue chronicles or discussion forums to survey travelogues, well before taking their journey. While reading travelogues may be fun and enjoyable for some, digging into numerous travelogues to find useful trip information is a tedious and boring task for many. Thus, there is a need for developing automatic travelogue mining techniques to convey useful information in a travelogue to its readers in a more effective way.

Most of the work (travel pattern [1], ROI [2], landmark recognition [3]) used user contributed photo on Flickr to mine knowledge from images. Only less research endeavours have been committed to learning mining from travelogues. Hao et al. [4] create an area diagram approach, which first mines area delegate terms from travelogues and after that utilizations such terms to recover Web pictures. The scholarly terms and recovered pictures are at long last exhibited to give an educational diagram of a given area. In [5], the creator presented the idea of worldwide and neighbourhood themes where general things are gathered into worldwide points while terms identified with nearby occasions are bunched into neighbourhood subjects. This paper investigates the travelogues and discovers the points on travelogues. It mines delegate labels, appropriation of cost and time of every theme. In light of these components, it semantically positions the travelogues. Envisioning an unadulterated literary travelogue with area based labels and pictures make it helpful for peruses to comprehend the main content of the travelogues.

This paper proposes a Travelogue RankTagViz approach that semantically ranks and visualizes user uploaded travelogues. The division of the paper is described as follows. Section 2 provides the related work of travelogue mining and visualization. Section 3 explains proposed travelogue RankTagViz approach and Section 4 discusses the performance analysis of proposed results. Finally Section 5 summarizes the work which has been done.

2. Related Work

This section mainly introduces three aspects of related works (i) travelogues mining; (ii) semantic ranking; (iii) visualization of travelogues.

User created travelogues give rich data. Kurashima et al. [8] removed run of the mill client's movement arrangements as indicated by passages, related with sight and sound data of the courses. Travel Scope system is designed by Hao et al. [6] to mine client

contributed travelogues and photos for virtual visits. It can prescribe prevalent spots with far reaching viewpoints as well as give delegate perspectives of points of interest. It gives three utilitarian perspectives to various levels of geo-granularity: suggestion see for a locale, viewpoint see for an area, and milestone see for a point of interest.

In [4] a framework was proposed to generate a representative and comprehensive location overview by mining textual travelogues. Given a set of travelogues collected from the Web shared by previous travellers, it is possible to extract information and generate location-representative tags. These representative tags can be combined with the location name to retrieve images related to a location from the Web, providing to the user a visual and textual description of a given destination. Ye et al. [7] built up a travelogue benefit that finds and passes on different travelogue digests, in the type of subject areas, geological extension, voyaging direction and area piece, to clients. In [9] a travelogue direction framework was suggested that naturally perceive and rank the historic points for voyagers. Travel Guide to distinguish and rank milestones for any area indicated by the explorers. Du and Hai [10] proposed a strategy for estimating site page closeness in light of formal ideas investigation (FCA). It likewise finds the semantic positioning of the site page.

Ghose et al. [11] create a positioning framework that prescribes items that give, by and large, the best an incentive for the shopper's cash. It utilized content mining together with swarm sourcing techniques to gauge requests for inns. Liu et al. [12] proposed to rank the picture labels as per their pertinence as for the related pictures by label closeness and picture comparability in an arbitrary walk show. Fan et al. [13] propose a novel substance based social label positioning plan, meaning to prescribe the semantic labels that the portrayals may not contain. The plan right off the bat secures the quantized semantic connections between words with experimental strategies, at that point develops the weighted tag-digraph in view of the portrayals and gained quantized semantics, lastly plays out an adjusted chart based positioning calculation to refine the score of every competitor tag for the suggestion. This paper utilizes semantic rank component that rank travelogues in view of area cost, going by time, going by season and labels.

In [8], imagine the exercises of visitors at touring spots. The framework acquires affiliation administers between areas, eras, and sorts of encounters in view of blog sections. Xue et al. [14] depict an approach for outwardly condensing a milestone by suggesting pictures with various perspectives. It displays a picture's perspective utilizing a 4-D perspective vector, which portrays perspective in level, vertical, scale and introduction angles. Ren et al. [15] create visual rundown for POI. (Purpose Of-Interest) through LOI (Location-Of-Interest) mining. It used to tag and area based separating to develop applicant POI information.

Rudinac et al. [16] exhibited a multimodal way to deal with visual rundown of geographic territories utilizing group contributed pictures. The approach makes utilization of the visual substance of the pictures, related comments (i.e., title, depiction and labels). Lu et al. ([17]) have proposed models for delineating travelogues by separating area particular themes from travelogues and finding pertinent pictures from labelled picture vault. The Tag Suggestr [18] framework in light of a blend of visual and content highlights required a client to clarify a photograph with an underlying arrangement of labels. The underlying labels were utilized to recover extra labels from related photographs that had a portion of the underlying labels in their label records. String et al. [19] proposed a system that finds area agent labels from travelogues and afterward select pertinent and delegate photographs to picture these labels

3. Travelogue RankTagViz

This paper proposes a Travelogue RankTagViz system which automatically mining, ranking and visualizing the travelogues. Figure 1 shows the overview of system architecture. The main components of the system model are as in the following:

3.1. Data Collection:

In this process, the user generated travelogues are collected from the web. Getting the information in travelogues is a non-inconsequential errand, on the grounds that most travelogues are unstructured and contains much clamour. In particular, in a normal travelogue, the area depictions are interwoven with closely-held convictions or other regular points, for example, convenience and transportation. So, collected travelogues are must be pre-processed for further use.

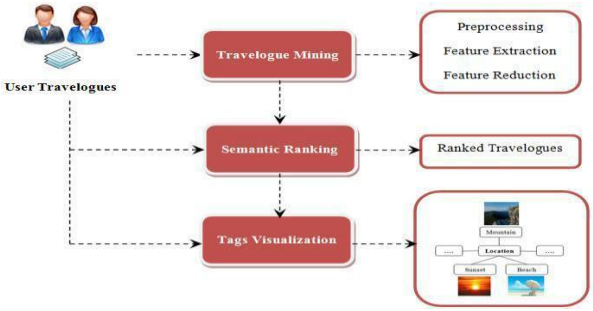


Fig. 1. System Architecture

3.2. Pre-processing:

In this process, the collected user generated travelogues are pre-processed. Stop words (a, an, are, as, it, of), i.e. words thought not to pass on any significance, are expelled from the content. Stop words are dialect particular useful words which convey no data. It might be of the accompanying kinds, for example, pronouns, relational words, conjunctions. The rationale that stop-words ought to be expelled from a content is that they influence the content to look heavier and less essential for investigators. Expelling stop words lessens the dimensionality of term space.

3.3. Feature Extraction:

In this process, the important and meaningful tags are extracted from the pre-processed travelogues. Grammatical form (POS) labelling is the way toward appointing a word to its syntactic classification, keeping in mind the end goal to comprehend its part inside the sentence. Conventional parts of discourse are things, verbs, modifiers, conjunctions, and so forth. Initially the noun words are selected as term or tag.

3.4. Feature Dimension Reduction:

In this process, extracted features (term or tag) are reduced. Each travelogue contains hundreds of terms which are considered as features or tags. All the feature value need not be available for all the travelogues as they differ from each other. A dimensionality reduction reduces the feature space to a much smaller number of dimensions, so that computations become tractable

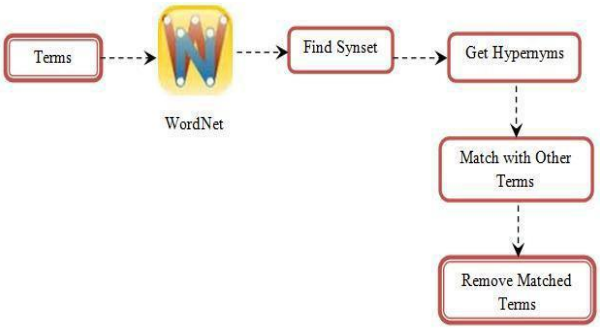


Fig. 2: Proposed architecture

In this paper the features are removed based on tfidf and WordNet. Figure 2 shows the feature dimension reduction algorithm. In the first step, Term Frequency Inverse Document Frequency (TF-IDF) strategy is utilized to get the score of each tag. Label score is utilized to mirror the significance of a tag to the point. The base score esteem labels are expelled from the term list. In the second step, the rest of the terms are lessened in view of the WordNet.

WordNet comprises of an arrangement of equivalent words "synsets". A synset signifies an idea or a feeling of a gathering of terms. Synsets give diverse semantic connections, for example, synonymy, antonym, hypernym, hyponymy, meronym and holonym.

In this paper the features are removed based on tfidf and WordNet. Figure 2 shows the feature dimension reduction algorithm. In the first step, Term Frequency-Inverse Document Frequency (TF-IDF) technique is utilized to get the score of each tag. Label score is utilized to mirror the significance of a tag to the point. The minimum score value tags are removed from the term list. In the second step, the remaining terms are reduced based on the WordNet.

Figure 3 shows the flow of term reduction based on wordnet. The wordNet process each terms and find the synset and hypernyms. The hypernyms are matched with other terms. If the matched terms are found then the particular term will be removed.

Tag Visualization In this process, the location based tags and images are extracted from the travelogues for visualization. Based on the ranked travelogues both tags and images are shortlisted for visualization. GraphViz [23] tool is used to visualize the tags and images of particular location.

4. Experimental Results

This section explains the performance evaluation of proposed approach. Travelogue RankTagViz is implemented. using Java (version 1.7), and the experiments are performed on a Intel(R) Pentium machine with a speed 4.13 GHz and 4.0 GB RAM using Windows 10 64-bit Operating System. This work collects travelogues from IndiaTravelogue [20], holidify [21] and Inditales [22] to form a corpus covering 100 unique locations in India and is Pre-processed by removing stop word and stemming. After preprocessing list of tags and POIs are extracted.

Feature Dimension Reduction Algorithm

Step 1: Reduction based on tfidf

Compute Term Frequency tf(i,j)

i=1,2,3,..n (No of Travelogues)

j=1,2,3...m (No of Terms)

Compute Document Frequency df(i)

Compute tfidf score

$$tfidf(i,j) = tf(i,j) * \log\left(\frac{N}{df(j)}\right)$$

Remove minimum score terms

Step 2: Reduction based on WordNet

Terms T_i (i=1,2,3,...m)

Add all the terms to Queue Q.

While (Q is not empty)

t1=Q.front()

Find the synset S for t1.

For i=1 to m

Map each term ti with synset S

If (not matched)

Term t1 is added to result list

End For

Q.pop

End While

Table 1: Sample Location With States

Location	State
Varanasi	Uttar Pradesh
Ladakh	Jammu & Kashmir
Bangalore	Karnataka
Kanyakumari	Tamil Nadu
Alleppey	Kerala
Pune	Maharashtra
Gangtok	Sikkim
Jaipur	Rajasthan

Word Score can be calculated as total number of words matched with reduced feature divided by the total number of words in travelogue. Sentence Score can be calculated as total number of sentence contains reduced feature divided by total number of sentence in travelogue.

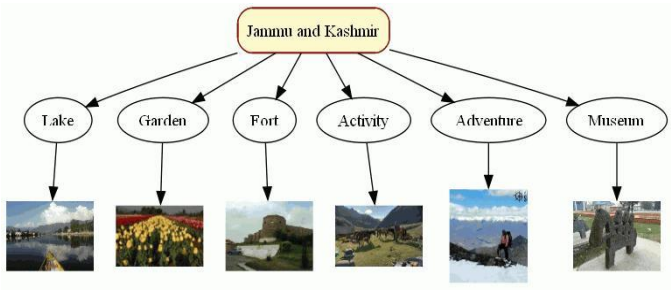


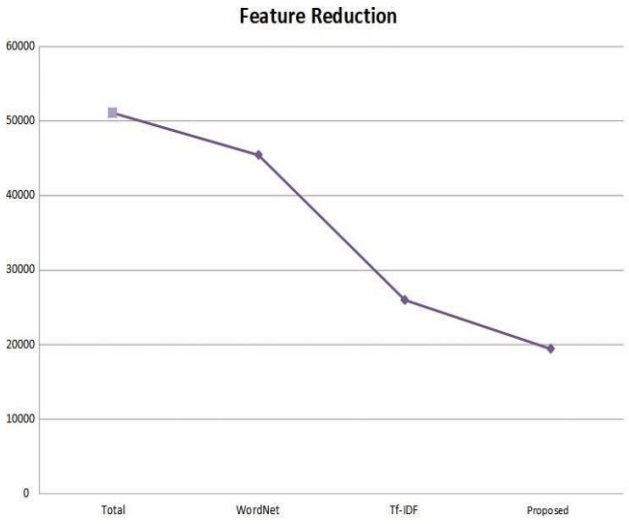
Fig. 3: The tag and image visualization of Jammu and Kashmir

Table 2 shows the feature reduction carried out on the travelogues by both the WordNet and TF-IDF. The results are convincing but the proposed work shows much better results while comparing.

Table 2: Features Selected By Existing And Proposed Methods

Total No of Feature	**51100**
WordNet	45449
TF-IDF	25992
Proposed	19454

The graphical representation of the acquired results is shown as follows,



5. Conclusion

Travelogues contain abundant location-representative knowledge, which is informative for other tourists, but difficult to extract and summarize manually. This paper proposed a framework Travelogue RankTagViz for mining knowledge from user generated travelogues. With this work, it effectively, reduce list of features using WordNet, rank user travelogues based on semantic concepts, and visualize location with tags and images. The proposed work was evaluated based on collected India travelogues, showing promising results on the above tasks. This proposal helps in detecting safe places to travel and also assist in exploring unknown destinations while on the go.

References

[1] Zheng, Y. T., Zha, Z. J., & Chua, T. S. (2012). "Mining travel patterns from geotagged photos". ACM Transactions on Intelligent Systems and Technology, 3(3), 1-18
[2] G. Cai, C. Hio, L. Bermingham, L. Kyungmi and I. Lee. (2014). "Mining Frequent Trajectory Patterns and Regions-of-Interest from Flickr Photos" . In Proc. of HICSS, 2014
[3] Yi-Hau Liu, Shian-Hua Lin, Chun-Ku Lai, Chun-Che Huang, Cheng-Yu Lu. (2016). "Mining Crowdsourcing Photos for Recognizing Landmark Areas", IMIS 2016: 12-19
[4] Q. Hao, R. Cai, X.-J.Wang, J.-M. Yang, Y. Pang, and L. Zhang.(2009). "Generating location overviews with images and tags by mining user-generated travelogues". In Proc. ACM Multimedia, 2009
[5] Q. Hao, R. Cai , C. Wang , R. Xiao , J.-M. Yang, Y. Pang, L. Zhang (2010). "Equip tourists with knowledge mined from travelogues". In: Proceedings of the 19th international conference on World wide web.ACM, pp 401–410
[6] Q. Hao, R. Cai, J. M. Yang et al. (2009). "TravelScope: standing on the shoulders of dedicated travelers," in Proceedings of the 17th ACM International Conference on
[7] Multimedia, MM'09, with Co-located Workshops and Symposiums, pp. 1021–1022, October 2009.
[8] M. Ye, R. Xiao, W. C. Lee, and X. Xie, (2011). "On theme location discovery for travelogue services," in Proceedings of the 34th international ACM SIGIR conference on Research and development in Information, pp. 465–474, 2011
[9] T. Kurashima, T. Tezuka, and K. Tanaka,(2006). "Mining and visualizing local experiences from blog entries," in Database and Expert Systems Applications. Springer, 2006, pp. 213–222.
[10] Y. Gao et al. (2010). "W2go: a travel guidance system by automatic landmark ranking". In ACM MM, 2010.
[11] Y. Du, Y. Hai. (2013)."Semantic ranking of web pages based on formal concept analysis" Journal of Systems and Software. 2013; 86(1): 187-97.
[12] A. Ghose, P.G. Ipeirotis, and L. Beibei, (2012). "Designing ranking systems for hotels on travel search engines by mining user-generated and crowdsourced content", Marketing Science, 31, 3, 493-520.
[13] D. Liu, X.-S. Hua, L. Yang, M. Wang, and H.-J. Zhang,(2009). "Tag ranking," in Proc. ACM Conf. World Wide Web, 2009, pp. 351–360.
[14] M. Fan, Q. Zhou, and T. F. Zheng,(2012). "Content-based semantic tag ranking for recommendation," in WI-IAT, 2012
[15] Y. Xue, , X.Qian. (2012) "Visual summarisation of Landmarks via viewpoint modeling", Proc, ICIP, 2012, pp. 2873–2876
[16] Y. Ren,X.Qian,andS.Jiang, (2015) "Visual Summarization for Place-of-Interest by Social Contextual Constrained Geo-clustering," in Proc. MMSP'15, 2015.
[17] S. Rudinac, A. Hanjalic, M.A. Larson, (2013) "Generating Visual Summaries of Geographic Areas Using Community - Contributed Images", IEEE Transactions on Multimedia, 15(4), pp. 921-932, 2013.
[18] X. Lu, Y. Pang Q. Hao and L. Zhang, (2009)."Visualizing textual travelogue with location-relevant images. In Proceedings of the 2009 International Workshop on Location Based Social Networks(2009), ACM, pp. 65–68
[19] O. Kucuktunc, S.G. Sevil, A. Tosun, H. Zitouni, P. Duygulu, and F. Can. (2008) "Tag suggestr: Automatic photo tag expansion using visual information for photo sharing websites". In Proc. Semantic and Digital Media Technologies: Semantic Multimedia , pages 61–73, 2008.
[20] Y. Pang, Q. Hao, Y. Yuan, T. Hu, R. Cai, and L. Zhang, (2011), "Summarizing tourist destinations by mining user-generated travelogues and photos," Comput. Vis. Image Understand. , vol. 115, no. 3, pp. 352–363, Mar. 2011.
[21] https://www.indiatravelogue.com/
[22] https://www.holidify.com/
[23] https://www.inditales.com/